PURDUE UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING
ELECTRONIC SYSTEMS RESEARCH LABORATORY

SEMI-ANNUAL REPORT OF RESEARCH

PERFORMED UNDER GRANT NsG - 553

January 1, 1965 through June 30, 1965

SCIENCE TECHNOLOGY AGRICULTURE

GPO PRICE $ ______

CFSTI PRICE(S) $ ______

Hard copy (HC) 2.00

Microfiche (MF) .50

ff 653 July 65

Lafayette, Indiana

N 66-23832

(ACCESSION NUMBER) (THRU)

47 1

(PAGES) (CODE)

CR-71865 07

(NASA CR OR TMX OR AD NUMBER) (CATEGORY)

FACILITY FORM 602

FOREWORD

This report constitutes a semi-annual review of the research supported in whole or in part under NASA Grant NsG-553 for the period January 1, 1965 - June 30, 1965. For an overview of the work reported here, reference must be made to the Annual Report of Research performed under this grant by the Electronic Systems Research Laboratory of Purdue University dated January, 1965.

This current summary of work progress over the six months' period through June, 1965, is not only a continuation of the projects described in the above-mentioned Annual Report but is also an attempt to give a fairly comprehensive view of the areas under research and, therefore, covers quite detailed descriptions of the projects.

TABLE OF CONTENTS

PROJECT PERSONNEL

Professorial Staff

D. R. Anderson	J. C. Lindenlaub
J. C. Hancock	R. E. Totty

P. A. Wintz

Instructors and Graduate Assistants

C. C. Bailey	D. F. Mix
R. W. Chang	E. A. Patrick
W. D. Gregg	E. A. Quincy
E. J. Luecke	M. D. Shapiro

I. LEARNING SYSTEMS

A. SIMULATED LEARNING SYSTEMS

J. C. Hancock

D. F. Mix

In the Annual Report of Research, January, 1965, a comparison among four learning systems was made. These were 1) with teacher, 2) averaging over all partitions, 3) decision-directed measurements, and 4) the iterative procedure. In each system, the a priori density function for the unknown parameter is modified by conditioning on the past samples; i.e., the decision is made by computing $p(x|w_1, x_1, ---; x_k)$ where

$$p(x|w_1, x_1, --, x_k) = \int p(x|w_1, \theta) p(\theta|x_1, --; x_k)\, d\theta \tag{1}$$

and computing $p(x|w_2, x_1, x_2, --x_k)$ where

$$p(x|w_2, x_1, --, x_k) = \int p(x|w_2, \emptyset) p(\emptyset|x_1, --, x_k) d\emptyset \tag{2}$$

and comparing the ratio of (1) over (2) to a fixed threshold.

A new "learning system" has been developed for learning the mean value of x where only the first two moments of θ and $\emptyset$ are needed--not the a priori density functions $p(\theta)$ and $p(\emptyset)$. The procedure is as follows:

Assume we know θ_o and $\emptyset_o$, the mean values of random variables θ and $\emptyset$, respectively, along with σ_{Lo}^2, the variance of both θ and $\emptyset$, and σ_n^2, the variance of x. After receipt of the first (unclassified) sample x_1, new estimates of θ and $\emptyset$ are calculated by

$$\theta_1 = \frac{\sigma_n^2 \theta_o + \sigma_{Lo}^2 x_1}{\sigma_n^2 + \sigma_{Lo}^2} P(w_1|x_1) + \theta_o[1 - P(w_1|x_1)] \tag{3}$$

$$\emptyset_1 = \frac{\sigma_n^2 \emptyset_o + \sigma_{Lo}^2 x_1}{\sigma_n^2 + \sigma_{Lo}^2} P(w_2|x_1) + \emptyset_o[1 - P(w_2|x_1)] \tag{4}$$

where $P(w_1|x_1)$ is the probability that x_1 is an element of class w_1, given the value of x_1. This probability is calculated by

$$P(w_1|x_1) = \frac{p(x_1|w_1, \theta_o)P(w_1)}{p(x_1|w_1, \theta_o)P(w_1) + p(x_1|w_2, \emptyset_o)P(w_2)} \tag{5}$$

Since the functions $p(x|w_1)$ and $p(x|w_2)$ are known except for the unknown parameters θ and $\emptyset$, the "best guess" θ_o and $\emptyset_o$ are used in (5).

The variance σ_{Lo}^2 is now modified by

$$\sigma_{L1}^2 = \frac{\sigma_{Lo}^2 \sigma_n^2}{\sigma_{Lo}^2 + \sigma_n^2} \tag{6}$$

and upon receipt of the second sample x_2, new estimates are calculated by

$$\theta_2 = \frac{\sigma_n^2 \theta_1 + \sigma_{L1}^2 x_2}{\sigma_n^2 + \sigma_{L1}^2} P(w_1|x_2) + \theta_1[1 - P(w_1|x_2)] \tag{7}$$

$$\emptyset_2 = \frac{\sigma_n^2 \emptyset_1 + \sigma_{L1}^2 x_2}{\sigma_n^2 + \sigma_{L1}^2} P(w_2|x_2) + \emptyset_1[1 - P(w_2|x_2)] \tag{8}$$

where $P(w_1|x_2)$ and $P(w_2|x_2)$ are calculated, as in (5), by using the best available guess θ_1 and $\emptyset_1$ for the unknown parameters.

Note that this procedure may be extended to the multi-dimensional case, and also is easily extended to more than two classes. The result of long computer runs (10,000 samples each) is shown in Table 1, where there are three classes. Figure 1 shows probability of error for the binary case where $\theta - \emptyset = 4\ \sigma_n$. This probability of error curve compares favorably with the iterative procedure* introduced by Fralick[1], yet the complexity is greatly reduced. By the statement that this system compares favorably with the iterative procedure, we mean that the results are indistinguishable. The computer time required to obtain Fig. 1 was 0.6 minutes, compared to 13 minutes for the corresponding graph for the iterative procedure.

REFERENCES

1. Fralick, S. C., "The Synthesis of Machines which Learn Without a Teacher," Tech. Report No. 6103-3, Stanford Electronics Laboratories, April, 1964.

*See Fig. 4, Page 4, in the Annual Report of Research Performed under Grant NsG-553, January, 1965.

Table I

Initial estimate	Final estimate	True value
1.0	1.98	2.0
3.0	5.52	5.0
10.0	8.30	7.0
2.5	1.02	1.0
3.5	4.62	5.0
5.5	8.75	9.0

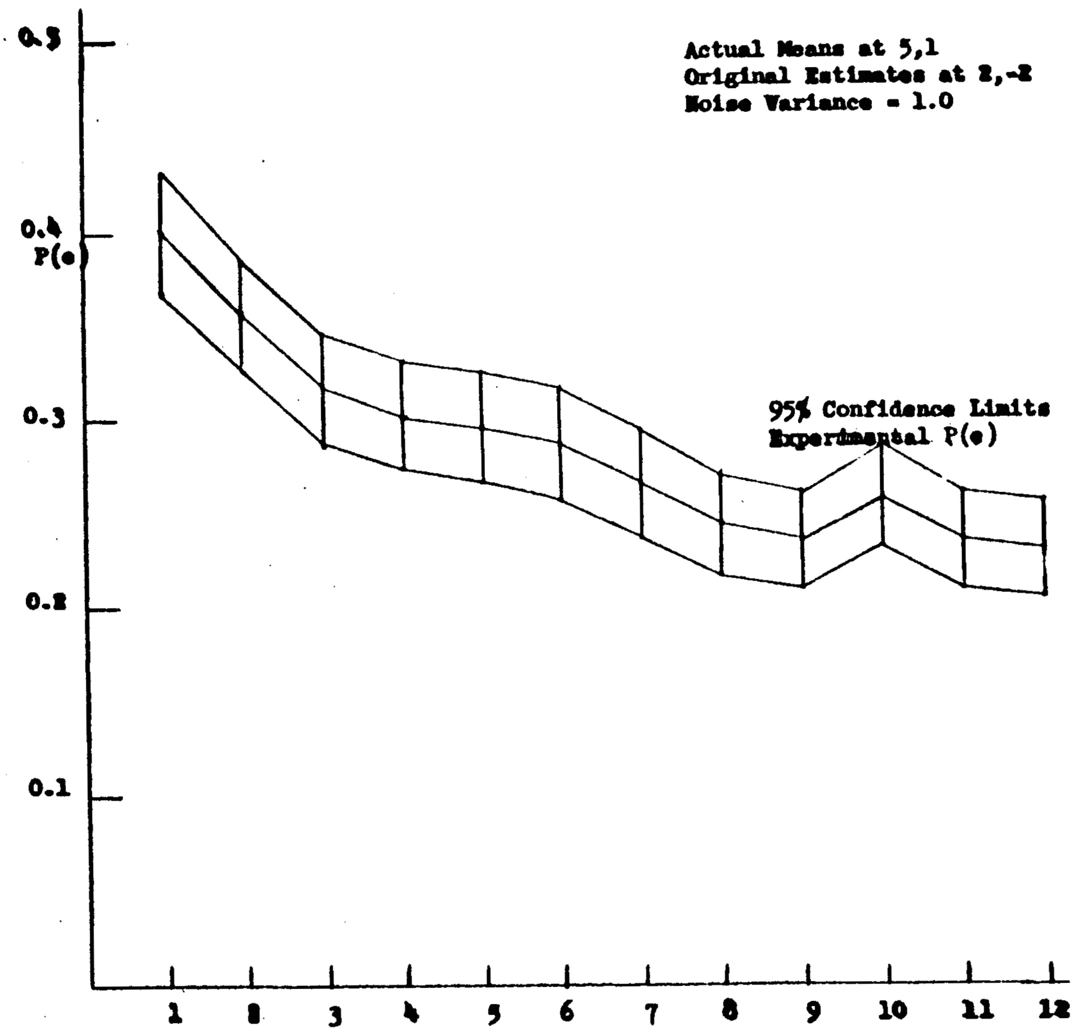

Fig. 1 Probability of Error Versus Number of Samples for the Simplified Iterative Procedure. Learning Variance = 1.0.

B. LEARNING PROBABILITY SPACES FOR CLASSIFICATION AND RECOGNITION OF PATTERNS WITH OR WITHOUT TEACHER

J. C. Hancock

E. A. Patrick

In the previous Annual Report, an approach using a "Fixed Bin" Model was introduced for learning probability spaces with or without supervision, with the objective of obtaining a system which minimizes conditional probability of error. Using this fixed bin model, the optimum system computes the conditional probability of the vectors $\underline{P}^i$, $i = 1,2,\ldots,m$, where $\underline{P}^i$ characterizes the conditional probability distribution function for the i^{th} class. By assuming that vectors $\underline{P}^i$ and $\underline{P}^j$, $i \neq j$, are statistically independent when conditioned on past samples, an iterative solution for $P(\underline{P}^i|\{X_s\}_n)$ was obtained in terms of $P(\underline{P}^i|\{X_s\}_{n-1})$. In general, this is suboptimum[1] since it is not, in general, true that $P(\underline{P}^i,\underline{P}^j|\{X_s\}_n) = P(\underline{P}^i|\{X_s\}_n)\ P(\underline{P}^j|\{X_s\}_n)$.

We have developed the optimum iterative solution for this nonsupervisory problem by approaching it as a mixture[2] problem, and correctly applying Bayes Theorem. We first establish that the class of conditional distribution functions, $\mathcal{F} = \{F(X|\omega_i)\}_m$, when forming a mixture, is identifiable[2,3].

A mixture of conditional distribution functions $\{F(X|\omega_i)\}_m$ is given in terms of mixing parameters $\{P(\omega_i)\}_m$ as follows:

$$F(X) = \sum_{i=1}^{m} P(\omega_i)\ F(X|\omega_i) \tag{1}$$

In Eq. (1), $F(X)$ is the mixing cumulative distribution function and is an identifiable mixture if for any other $\{\overline{F}(X|\omega_i)\}_m \in \mathcal{F}$ and $\{\overline{P}(\omega_i)\}_m$, then

$$F(X) = \sum_{i=1}^{m} \overline{P}(\omega_i)\ \overline{F}(X|\omega_i)$$

if and only if

$$\overline{P}(\omega_i) = P(\omega_i),\ \overline{F}(X|\omega_i) = F(X|\omega_i),\ I = 1,2,\ldots,m$$

We have considered those classes of c.d.f.'s which can be shown to be identifiable, and have been able to take into account such a priori information that the c.d.f.'s differ only by a translational parameter, are symmetrical, etc.

To see how fundamental the mixture approach is, let B_i be a vector of parameters characterizing the c.d.f. $F(X|\omega_i)$. For example, if $F(X|\omega_i)$ is known gaussian, then $B_i = (m_i, \sigma_i)$. If $F(X|\omega_i)$ is completely unknown and the fixed bin model is used, then $B_i = (p_1^{\,i}, \ldots, p_r^{\,i}) = \underline{P}^i$, the vector of bin probabilities. We also define the vector of mixing parameters, $B_{m+1} = (P(\omega_1), \ldots, P(\omega_m))$ and a vector B:

$$B = B_1 U B_2 U, \ldots, U B_m U B_{m+1} \tag{2}$$

The optimum system which minimizes conditional probability of error must compute $F(B|\{X_s\}_n)$. By Bayes Theorem

$$f(B|\{X_s\}_n) = \frac{f(\{X_s\}_n|B)f(B)}{f(\{X_s\}_n)} \tag{3}$$

Assuming the vector samples $X_1, X_2, \ldots, X_n$ are conditionally independent, we obtain

$$f(\{X_s\}_n|B) = \prod_{s=1}^{n} f(X_s|B) \tag{4}$$

From Eqs. (1, 2, 3, and 4) we obtain

$$f(B|\{X_s\}_n) = \frac{\left[\prod_{s=1}^{n} \sum_{i=1}^{m} P(\omega_i) f(X_s|\omega_i, B_i)\right] f(B)}{f(\{X_s\}_n)} \tag{5}$$

The optimum iterative solution corresponding to Eq. (5) is

$$f(B|\{X_s\}_n) = \frac{\left[\sum_{i=1}^{m} P(\omega_i) f(X_n|\omega_i, B_i)\right] f(B|\{X_s\}_{n-1})}{f(X_n|\{X_s\}_{n-1})} \tag{6}$$

It is very important to note that the iterative solution, Eq. (6), requires that the conditional joint probability density be used in the iteration.

For the last three years[4,5,6] several researchers were not able to obtain the very important result given by Eq. (6). Fralick[6], for example, incorrectly assumed that

$F(\theta_i,\theta_j|\{X_s\}_n) = F(\theta_i|\{X_s\}_n)F(\theta_j|\{X_s\}_n)$ where θ_i and θ_j are parameters characterizing the i^{th} class and j^{th} class, respectively. Such an assumption leads, in general, to extremely suboptimum systems.

We conclude by showing that an exponentially growing solution for the binary case[5] is a special case of Eq. (6).

Let X_{n+1} be the $n+1^{st}$ sample from a mixture of continuous distribution functions $F(X_{n+1}|\omega_i^{n+1},\theta_i)$ each depending on a single translational parameter θ_i. The i^{th} class, active on the $n+1^{st}$ sample, is denoted by ω_i^{n+1}, and it is assumed that $P(\omega_i^{n+1}) = P(\omega_i)$, $i=1,2,\ldots,m$, i.e., the class probabilities, are known.

The a posteriori probability of the event (X_{n+1},ω_i^{n+1}) was computed by Abramson[4], later by Daly[5], by partitioning the sequence of samples $X_1\ldots X_n$ into the mutually exclusive and exhaustive ways that they can occur. If π_r^n denotes the r^{th} of the m^n possible partitions, Abramson and Daly noted that

$$f(X_{n+1},\omega_i^{n+1}|\{X_s\}_n) = \sum_{r=1}^{m^n} f(X_{n+1},\omega_i^{n+1}|\pi_r^n,\{X_s\}_n)P(\pi_r^n|\{X_s\}_n) \tag{7}$$

$$= \sum_{r=1}^{m^n} f(X_{n+1}|\omega_i^{n+1},\pi_r^n,\{X_s\}_n)P(\omega_i^{n+1}|\pi_r^n,\{X_s\}_n)P(\pi_r^n|\{X_s\}_n) \tag{8}$$

$$= P(\omega_i) \sum_{r=1}^{m^n} f(X_{n+1}|\omega_i^{n+1},\pi_r^n,\{X_s\}_n)P(\pi_r^n|\{X_s\}_n) \tag{9}$$

$$= P(\omega_i) \sum_{r=1}^{m^n} \int f(X_{n+1}|\omega_i^{n+1},\theta_i)f(\theta_i|\pi_r^n,\omega_i^{n+1},\{X_s\}_n)d\theta_i P(\pi_r^n|\{X_s\}_n) \tag{10}$$

where $P(\omega_i)$, $i=1,s,\ldots,m$ are assumed known.

Fralick[6], looking for a simplification of Eq. (10), obtained an iterative form assuming that, if θ_i is a parameter of $f(X|\theta_i,\omega_i)$ and θ_j a parameter of $f(X|\theta_j,\omega_j)$, then $f(\theta_i|\{X_s\}_n,\theta_j) = f(\theta_i|\{X_s\}_n)$. Fralick's result is, in general, suboptimum since, in general, $f(\theta_i|\{X_s\}_n,\theta_j) \neq f(\theta_i|\{X_s\}_n)$ $j\neq i$. This condition is true when θ_j is known, which for $m = 2$, $\theta_i = \theta_1$, $\theta_j = \theta_2$, corresponds to the binary on-off case without supervision.

We show here that the desired a posteriori probability is either of the growing form or equivalently an iterative form for the joint a posteriori probability of $\theta_1 \cdots \theta_m$, where the marginal a posteriori probability of θ_i is obtained by integrating the joint density to get the marginal density. To show this result, we concentrate on $f(\theta_i|\pi_r^n,\omega_i^{n+1},\{X_s\}_n)$ in Eq. (10). Applying Bayes theorem to $f(\theta_i|\pi_r^n,\omega_i^{n+1},\{X_s\}_n)$ and noting that ω_i^{n+1} can be dropped from the expression gives

$$f(\theta_i|\pi_r^n,\{X_s\}_n) = \frac{f(X_n|\pi_r^n,\{X_s\}_{n-1},\theta_i)\; f(\theta_i|\pi_r^n,\{X_s\}_{n-1})}{f(X_n|\{X_s\}_{n-1},\pi_r^n)} \tag{11}$$

Substituting Eq. (11) in Eq. (10) gives, after some calculations,

$$P(\omega_i) \int f(X_{n+1}|\omega_i^{n+1},\theta_i)d\theta_i \sum_{r=1}^{m^{n-1}} \frac{\sum_{v=1}^{m} f(X_n,\omega_v^n|\pi_r^{n-1},\{X_s\}_{n-1},\theta_i)}{f(X_n|\{X_s\}_{n-1},\pi_r^{n-1})} P(\pi_r^{n-1}|\{X_s\}_{n-1}) \tag{12}$$

$$f(\theta_i|\pi_r^{n-1},\{X_s\}_{n-1})$$

such that

$$f(\theta_i|\omega_i^{n+1},\{X_s\}_n) = f(\theta_i|\{X_s\}_n) =$$

$$\sum_{r=1}^{m^{n-1}} \frac{\sum_{v=1}^{m} f(X_n,\omega_v^n|\pi_r^{n-1},\{X_s\}_{n-1},\theta_i}{f(X_n|\{X_s\}_{n-1},\pi_r^{n-1})} P(\pi_r^{n-1}|\{X_s\}_{n-1})\; f(\theta_i|\pi_r^{n-1},\{X_s\}_{n-1}) = \tag{13}$$

$$\sum_{r=1}^{m^{n-1}} \frac{\sum_{j\neq i} \left[P(\omega_j)f(X_n|\pi_r^{n-1},\{X_s\}_{n-1},\omega_j^n,\theta_i)\right] + P(\omega_i)f(X_n|\pi_r^{n-1},\{X_s\}_{n-1},\theta_i,\omega_i^n)}{f(X_n|\{X_s\}_{n-1})} \tag{14}$$

$$P(\pi_r^{n-1}|\{X_s\}_{n-1})\; f(\theta_i|\pi_r^{n-1},\{X_s\}_{n-1}) =$$

$$\sum_{r=1}^{m^{n-1}} \frac{\sum_{j\neq i} \left[\; P(\omega_j)\int f(X_n|\omega_j^n,\theta_j)f(\theta_j|\pi_r^{n-1},\{X_s\}_{n-1},\theta_i)d\theta_j \;\right] + P(\omega_i)f(X_n|\theta_i,\omega_i^n)}{f(X_n|\{X_s\}_{n-1},\pi_r^{n-1})} \tag{15}$$

$$P(\pi_r^{n-1}|\{X_s\}_{n-1})f(\theta_i|\pi_r^{n-1}\{X_s\}_{n-1})$$

where we have assumed conditional independence and $f(X_n|\omega_v^n,\theta_v)$ known for $v = 1,2,\ldots,m$. Interchanging integration with respect to both summation signs in Eq. (15) gives

$$f(\theta_i|\{X_s\}_n) =$$

$$\int\cdots\int \prod_{j\neq i} d\theta_j \sum_{r=1}^{m^{n-1}} \left[\sum_{j\neq i} P(\omega_j)f(X_n|\omega_j^n,\theta_j) + P(\omega_i)f(X_n|\omega_i^n,\theta_i)\right] P(\pi_r^{n-1}|\{X_s\}_{n-1}) \frac{f(\{\theta_j\}_{j\neq i},\theta_i|\{X_s\}_{n-1},\pi_r^{n-1})}{f(X_n|\{X_s\}_{n-1},\pi_r^{n-1})} \tag{16}$$

The outer summation is removed by the inverse operation to that used in Eq. (7) to obtain

$$f(\theta_i|\{X_s\}_n) =$$

$$\int\cdots\int \prod_{j\neq i} d\theta_j \left[\sum_{j\neq i} P(\omega_j)f(X_n|\omega_j^n,\theta_j) + P(\omega_i)f(X_n|\omega_i^n,\theta_i)\right] \frac{f(\{\theta_j\}_{j\neq i},\theta_i|\{X_s\}_{n-1})}{f(X_n|\{X_s\}_{n-1})} \tag{17}$$

$$= \int\cdots\int \prod_{j\neq i} d\theta_j \frac{f(X_n|\{\theta_j\}_{j\neq i},\theta_i)\ f(\{\theta_j\}_{j\neq i},\theta_i|\{X_s\}_{n-1})}{f(X_n|\{X_s\}_{n-1})} \tag{18}$$

Thus an equivalent form of the optimum, growing solution is an iterative solution which computes the marginal a posteriori densities from the joint a posteriori probability density according to Eq. (18) where

$$f(X_n|\{\theta_j\}_{j\neq i},\theta_i) = \sum_{j\neq i} P(\omega_j)f(X_n|\omega_j^n,\theta_j) + P(\omega_i)f(X_n|\omega_i^n,\theta_i) \tag{19}$$

is a mixture of conditional c.d.f.'s, $f(X_n|\omega_v^n,\theta_v)$, and mixing parameters $P(\omega_v)$, $v=1,2,\ldots m$.

A basic approach to this nonsupervisory problem is thus to start with Eq. (18) using the mixture expression of Eq. (19) and compute

$$f(X_{n+1},\omega_i^{n+1}|\{X_s\}_n) = P(\omega_i) \int f(X_{n+1}|\omega_i^{n+1},\theta_i)f(\theta_i|\{X_s\}_n)d\theta_i \tag{20}$$

In general, $f(\theta_i|\{X_s\}_n)$ is computed in terms of the joint probability density, $f(\theta_1 \ldots \theta_m|\{X_s\}_{n-1})$, at the $n-1^{st}$ stage.

REFERENCES

1. Patrick, E.A., and J. C. Hancock,"The Nonsupervised Learning of Probability Spaces and Recognition of Patterns," IEEE International Convention, Information II, March, 1965.

2. Teicher, Henry, "On the Mixture of Distributions," Annals of Math. Stat., Vol. 34, No. 4, December, 1963.

3. Teicher, Henry, "Identifiability of Finite Mixtures," Annals of Math. Stat., Vol. 34, No. 4, December, 1963.

4. Abramson, N., "Learning in Pattern Recognition," ITT Federal Laboratories, First Technical Note, October, 1961.

5. Daly, R.F., "The Adaptive Binary-Detection Problem on the Real Line," Stanford Electronics Laboratory Report TR2003-3, February, 1962.

6. Fralick, S.C., "The Synthesis of Machines which Learn without a Teacher," Stanford Technical Report No. 61308-8, April, 1964.

C. SYNTHESIS OF OPTIMUM RECEIVERS FOR M-ARY CHANNELS WITH EXTENSIVE INTERSYMBOL INTERFERENCE

J. C. Hancock

R. W. Chang

This study is a continuation of the work reported in "Annual Report of Research, Grant NsG-553, Jan., 1965"[1,2].

Due to the ever-increasing data rate in high speed digital communications, one has to deal with intersymbol interference in certain circumstances between not just two adjacent time periods, but quite a number of adjacent time periods. New results in this study are:

1) An optimum receiver structure is obtained for the general case of m-ary channels with intersymbol interference between a large number of observations.

2) A new concept of a posteriori weighting matrix is introduced which holds for the general problem of observing a markov source through a noisy channel. A matrix chain is derived showing a procedure of modifying a conventional markov chain equation with the a posteriori observations.

3) A new concept of sufficient decision statistics is introduced. A theorem is derived which shows how to obtain optimum receiver structures that can be implemented in practice.

Although the discussions in this study are phrased in terms of communication channels, the methods and concepts are completely general and can be applied to other problems. Since a markov chain of order $L > 1$ can be reduced to a markov chain, it suffices to consider a markov chain.

Consider a digital communication system in which a sequence of independent m-ary signal digits $B_1, \ldots, B_n$ is transmitted with known a priori probability. Because of the high data rate, there is intersymbol interference between L time periods (for example, $L = 5$). Let $S_k(t)$ be the total received signal in the kth time period, and denote the classification of $S_k(t)$ by A_k. It can be shown that

A_1, $A_2, \ldots, A_n$ form a markov chain.

In this study, the waveform of each possible $S_k(t)$ is known. $S_k(t)$ is contaminated by noise $N_k(t)$. The receiver observes $X_k(t) = S_k(t)+N_k(t)$. Assume $X_k(t)$ is sampled, and let X_k, S_k, and N_k be the vectors of the samples. Although the noise is not necessarily gaussian, the probability density function of N_k is known. The samples in N_k need not be independent; it is only assumed that the random vectors $N_1, \ldots, N_n$ are independent.

To classify the m-ary signal digits $B_1, \ldots B_n$ with minimum probability of error after receiving $X_1, \ldots X_n$, the optimum receiver computes $P(\pi = \pi_j/V_n)$, $j = 1, \ldots m^n$, and accepts the hypothesis $\pi = \pi_i$ if $P(\pi = \pi_i/V_n) > P(\pi = \pi_j/V_n)$ for all $j \neq i$, where $\pi = (B_1, \ldots, B_n)$ and $V_n = (X_1, \ldots, X_n)$.

The optimum receiver constructed in this manner will compute m^n probabilities. Since m^n increases exponentially with n, it is practically impossible to implement such a receiver for even moderate values of n (e.g., $m = 2$, $n = 20$). To overcome this difficulty, it is important to observe that if the probability of error is to be very small, as it is in most practical systems, the a posteriori probability $P(\pi = \pi_T/V_n)$ corresponding to the true classification π_T must be close to unity for large values of n(e.g., $m = 2$, $n = 200$). This observation leads to the following theorem.

Theorem: Let π_T be the true classification of π. If $P(\pi = \pi_T/V_n) > \frac{1}{2}$, a condition usually satisfied in practice, then the optimum (minimum probability of error) receiver can be constructed from the following decision rule:

Compute the probability $P(A_k = j/V_n)$, $j = 1, \ldots, m^L$, and accept the hypothesis $A_k = i$ if $P(A_k = i/V_n) > P(A_k = j/V_n)$ for all $j \neq i$.

According to this theorem, $P(A_k = j/V_n)$, $k = 1, \ldots, n$, $j = 1, \ldots, m^L$ form a set of sufficient decision statistics for optimum decision under the condition $P(\pi = \pi_T/V_n) > \frac{1}{2}$. The importance of this theorem lies in the fact that it introduces the new concept of sufficient decision statistics. By this concept, one can reduce a multi-dimensional decision statistic (such as π of n dimension) to a decision statistic of lower dimension (such as A_k of one dimension), thus greatly simplifying the data processing procedure. As will be shown in the following, $P(A_k = j/V_n)$ can be computed with a fixed receiver structure which does not grow with n. This receiver structure is simple and can be implemented in practice.

Comparing $P(A_k = j/V_n)$ is the same as comparing $P(A_k = j/V_n)\, p(V_n)$. It can be shown that $P(A_k = j/V_n)\, p(V_n)$ can be broken into two terms. The first term is $P(A_k = j/V_k)\, p(V_k)$ in which the past and present observations $V_k = (X_1, \dots X_k)$ are utilized in classifying A_k. The second term is $p(U_k/A_k = j)$ in which the future observations $U_k = X_{k+1}, \dots, X_n$ are utilized to classify A_k. Thus, the future observations can be handled separately for any value of L. This conclusion is contrary to that in Gonsalves' report[3].

The first term $P(A_k = j/V_k)p(V_k)$ is now studied. Derivations are omitted. Only results are given. Define four matrices:

$$P(k) = \begin{bmatrix} P(A_k = 1/V_k)\, p(V_k) \\ P(A_k = 2/V_k)\, p(V_k) \\ \vdots \\ P(A_k = m^L/V_k)\, p(V_k) \end{bmatrix}, \qquad P_o = \begin{bmatrix} P(A_1 = 1) \\ P(A_1 = 2) \\ \vdots \\ P(A_1 = m^L) \end{bmatrix}$$

$$L(K) = \begin{bmatrix} P(X_k/A_k = 1) & o & \cdots & o \\ o & P(X_k/A_k = 2) & \cdots & o \\ \vdots & \vdots & \ddots & \vdots \\ o & o & \cdots & P(X_k/A_k = m^L) \end{bmatrix}$$

$$P(K,k-1) = \begin{bmatrix} P(A_k = 1/A_{k-1} = 1) & P(A_k = 1/A_{k-1} = 2). & \cdots & .P(A_k = 1/A_{k-1} = m^L) \\ P(A_k = 2/A_{k-1} = 1) & P(A_k = 2/A_{k-1} = 2). & \cdots & .P(A_k = 2/A_{k-1} = m^L) \\ \vdots & & \ddots & \\ P(A_k = m^L/A_{k-1} = 1) & P(A_k = m^L/A_{k-1} = 2). & \cdots & .P(A_k = m^L/A_{k-1} = m^L) \end{bmatrix}$$

We obtain

$$P(k) = L(k)\, P(k,k-1)\, P(k-1) \tag{1}$$

Note that Eq. (1) is in iterative form. Iterating Eq. (1) gives

$$P(k) = L(k)\, p(k,k-1)\, L(k-1)\, P(k-1,k-2) \ldots\ldots\ldots L(2)\, P(2,1).L(1)\, P_o \tag{2}$$

Equation (2) is an important result. The matrices $P(k,k-1)$, $P(k-1,k-2),\ldots,P(2,1)$, and P_o in Eq. (2) correspond to the transition matrices defined in the markov chain study. They provide the a priori information concerning the classification of A_k because their elements are given a priori probabilities. The other matrices $L(k)$, $L(k-1),\ldots,$ $L(2)$, and $L(1)$ in Eq. (2) can be termed the a posteriori weighting matrices. They provide the a posteriori information about the classification of A_k, as their elements are the likelihood functions computed from the observations $X_1,\ldots,$ X_k. The a priori and the a posteriori information can be handled separately and then combined as in Eq. (2). As far as the authors are aware, Eq. (2) has not appeared in the past. It is not possible to describe the related results and generalizations here; they will be included in a technical report (also in a paper which has been submitted to IEEE[4]).

The above considers the first term $P(A_k = j/V_k)\, p(V_k)$ of the decision statistic $P(A_k = j/V_n)\, p(V_n)$. The second term $p(U_k/A_k = j)$ of the decision statistic can be computed from the equation (derivations omitted)

$$Q(k) = P^T(k+1,k)\, L(k+1)\, P^T(k+2,k+1)\, L(k+2). \ldots P^T(n,n-1)L_n \tag{3}$$

where $P^T(k+1,k)$ is the transpose of $P(k+1,k)$ and

$$Q(k) = \begin{bmatrix} P(U_k/A_k = 1) \\ P(U_k/A_k = 2) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ P(U_k/A_k = m^L) \end{bmatrix} \quad \text{and} \quad L_n = \begin{bmatrix} p(X_n/A_n = 1) \\ p(X_n/A_n = 2) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ p(X_n/A_n = m^L) \end{bmatrix}$$

Equation (3) can also be written in iterative form as

$$Q(k) = P^T(k+1,k)\ L(k+1)\ Q(k+1) \tag{4}$$

Decision statistic $P(A_k = j/V_n)\ p(V_n)$ can be computed by combining $P(k)$ and $Q(k)$. Decision rule can then be applied to classify the observation $X_1, \ldots, X_n$. This completes the data processing procedure.

It can be seen from Eqs. (1) and (4) that only two matrices are involved in each iteration. Thus, the receiver structure is fixed and simple, and can be implemented in practice.

REFERENCES

1. Hancock, J. C., and Chang, R. W., "Unsupervised Learning Receivers for Binary Channels with Intersymbol Interference," Proc. IEEE Symp. on Signal Transmission and Processing, New York, N. Y., May 1965.

2. Hancock, J. C., and Chang, R. W., "Unsupervised Learning Receivers for Binary Channels with Intersymbol Interference", Electronic System Research Laboratory, Purdue University, Lafayette, Ind., Technical Report No. TR-EE 65-2, Jan., 1965.

3. Gonsalves, R. A., and Lob, W. H., "Maximum Likelihood Detection in a Binary Channel with Memory," Northeastern University, Scientific Report No. 4, July, 1963.

4. Chang, R. W., and Hancock, J. C., "Synthesis of Optimum Receivers for M-ary Channels with Extensive Intersymbol Interference," Submitted to IEEE Transaction on Information Theory.

D. COGNITIVE SIGNAL PROCESSING

J. C Hancock

W. D. Gregg

1. Review of the Research

This research is motivated by the vast requirements for processing the following classes of signals:

(a) Signals of electromagnetic origin; multiple category radar echos for discrimination of radar cross-section signatures (missile-decoy cross-sections, clear air turbulence cross sections, fine structure characterized surface cross-sections, etc.); signals arising in digital data links which have experienced random multiplicative fading and beam splitting (multipath) with phase dispersion.

(b) Signals of acoustic origin arising as a result of medium sounding for structure of object detection and acoustic cross-section discrimination as in seismic or submarine sounding or passive detection of submarines and other acoustic sources.

(c) Afferent signals of bioelectric origin which contain the features and signatures of a physical environment encoded by the sensory transducers.

The research is currently* concerned with the problem of optimum (defined below) signal processing when

(a) Statistical and deterministic features of the categories (disturbance and signal) are unknown.

(b) Classified or supervised time slots (classified learning sequences) are not available for a priori estimation of the above features.

The initial investigation has been restricted to the "low pass" two category or binary case which might be represented by any of the waveforms below,

*Purdue University, Electronics Systems Laboratory, Annual Report of Research, Jan., 1965, p. 15.

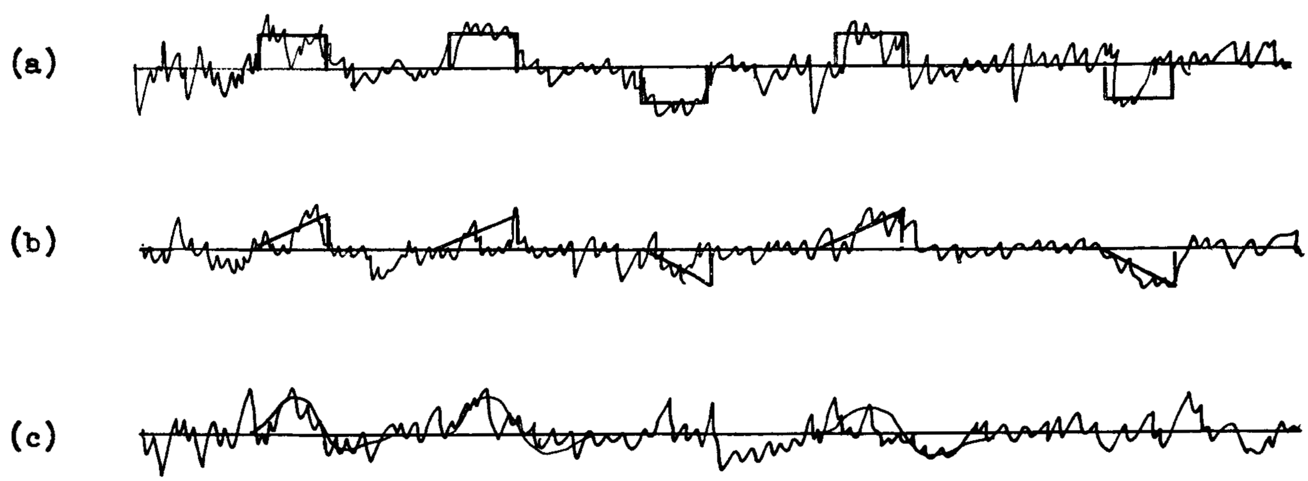

Fig. 1

2. Approach to the Problem

An attempt has been made to approach the problem by applying the fundamentals of mathematical statistics pertinent to discriminant function theory and generalized hypothesis testing. The initial assumptions are

(a) N discrete vector observations (sampled time slots) available for processing.

(b) V of the vectors are from ω_1 and W of the vectors are from ω_2 with $V + W = N$ and $\omega_1 \cup \omega_2 = \Omega = \mathcal{E}_n$

(c) The proportionality factors V and W are not known and the vector sequence $\underline{Z} = \{\underline{z}_n\}$ is not classified.

The principal difficulty in unsupervised or unclassified discrimination (dichotomization in the two category case) is due to the lack of a reference for "comparing" the $\underline{z}_i$, a situation that does not exist in the classical hyper-plane (Bayes discrete matched filter) case or the supervised case. Thus the "reference" must be generated from within the time series.

The strategy used to establish a "reference" from within the time series consists of testing successive data vectors against the preceding data vector using all prior data vectors for information content about the parameters of discrimination. The tests can be either parametric or distribtution free. The parametric tests are carried out via the generalized likelihood ratio test given by

$$\lambda_j(\underline{z}_i,\underline{z}_{i+1}|\underline{z}_{i-1}\cdots\underline{z}_1) = \frac{\underset{\omega_i\cap\omega_{i+1}}{\text{Max}}\ P_{2n}(\underline{z}_i,\underline{z}_{i+1}|\underline{z}_{i-1}\cdots\underline{z}_1)}{\underset{\Omega}{\text{Max}}\ P_{2n}(\underline{z}_i,\underline{z}_{i+1}|\underline{z}_{i-1}\cdots\underline{z}_1)} = \frac{\underset{\underline{\theta}_i,\Theta}{\text{Max}}\ L(\omega_1\cap\omega_2)}{\underset{\underline{\theta}_1,\underline{\theta}_2,\Theta}{\text{Max}}\ L(\Omega)} \qquad (\;)$$

$$i,i+1 = 1,2 \quad ; \quad j = 1,2,\ldots, M$$

The optimality of the generalized likelihood ratio test lies in the fact that if a uniformly most powerful (hence minimal probability of error) test exists, it is given by $\lambda_j(\underline{z}_i,\underline{z}_{i+1}|\underline{Z})$. A useful practical aspect is due to the fact that $-2 \log \lambda_j$ is asymptotically chi-squared. Hence successive isolations and separations of the data vectors occur as the null and alternative hypotheses H_o, H_1 are designated by the values of λ_j.

It is appropriate to point out that investigations and analyses of other classes of Bayes approaches other than the classical Bayes approach are considered for the following reasons.

(a) The "classical" Bayes approach requires the computation of

$$P(\underline{z}_{N+1}|\underline{Z};\omega_i) = \int_{R_{\{\theta\}}} P(\underline{z}_{N+1}|\{\theta\};\omega_i)P(\{\theta\}|\underline{Z};\omega_i)\ d\{\theta\} \qquad (2)$$

$$i = 1,2 \quad ; \quad \underline{Z} = \{\underline{z}_N\}$$

where $\{\theta\}$ is the multidimensional set of parameters of category i including the location parameter $\underline{\theta}$.

(b) For the classified, supervised, or the "learning with teacher" case, the posterior density kernel of the assumed random parameter $\{\theta\}$ is

$$h(\{\theta\}|\underline{Z};\omega_i) = h(\{\theta\})\ h(\underline{Z}|\{\theta\}\ ;\ \omega_i) \qquad (3)$$

where $h(\{\theta\})$ is the prior kernel of $\{\theta\}$ and $h(\underline{Z}|\{\theta\};\omega_i)$ is the kernel of the likelihood of $\underline{Z}$ given $\{\theta\}$. For the assumptions of a conjugate[1] prior kernel on $\{\theta\}$, the posterior kernel $h(\{\theta\}|\underline{Z};\omega_i)$ is of the same form (reproducibility), however the prior or learning sequence must be classified. The likelihood $P(\underline{z}_{N+1}|\underline{Z};\omega_i)$ need not have the same form as the posterior kernel.

(c) For the unsupervised, unclassified, or "without teacher" case about the prior sequence $\underline{Z}$, the "classical" Bayes form upon expansion of $P(\{\theta\}|\underline{Z})$ has yielded[5].

(1) $P(\{\theta\}|\underline{Z})$ containing 2^N terms as a result of a Bayes expansion conditioned upon all possible formats of the sequence $\underline{Z} = \{\underline{z}_N\}$ requiring 2^N computations or the equivalent of "2^N likelihood structures".

(2) $P(\{\theta\}|\underline{Z})$ expanded in a particular manner with particular assumptions and the introduction of a mixture expansion at a particular stage effecting a recursive form requiring computations over all possible $\{\theta\}$.

In both exponential and recursive forms, the final result is not unique and depends upon the order of intermediate assumptions and manipulations. Thus in addition to the less desirable properties of requiring 2^N computations or computations over all possible values of $\{\theta\}$ (completely different averages), the final form does not appear to be unique.

3. Physical Filter Interpretation of Cognition and Learning

It must be indicated that the assumption of N sampled time slots, $\underline{z}_j$, assumes time slot synchronization as well as a knowledge of the category duty cycle τ. Thus the argument might be posed that a knowledge of the duty cycle allows the selection of a time constant for a simple first order low-pass RC filter with continuous filtering yielding a degradation of only 1 db in the ratio of peak pulse power to mean square noise voltage due to mismatch. This would be true if the signal pulses were square and this fact was known; however, if the pulse shape is more complex, such as a sawtooth pulse of equivalent energy, the "optimum" time constant differs considerably from that for the square pulse (see Fig. 2 where ω_c is the 3 db filter radian frequency). Thus a selection of the "optimum" mis-matched or sub-optimum first order matched continuous filter on the basis of pulse duty cycle would result in an unnecessary degradation of SNR_o. This effect observed for the cases of rather simple low-pass signal structure in additive white noise certainly justifies the attempt to "learn" the category features, characteristics, or signatures.

4. Current Results and Conclusions.

A portion of the initial effort has been devoted to a study of the behavior of the power function, $\beta(\delta/\sqrt{\theta_3})$, for a given significance level of "one shot" vector sequence dichotomizations of waveform (a) on the basis of statistical feature dissimilarity about the location parameter $\underline{\theta}$ under the following conditions.

(a) Means, $\underline{\theta}_1$, $\underline{\theta}_2$ unknown, variance θ_3 known; Parametric (Gaussian)

(b) Means, $\underline{\theta}_1$, $\underline{\theta}_2$ unknown, variance θ_3 unknown; Parametric (Gaussian)

(c) Non-Parametric (Distribution Free); Sign Test.

This portion of the analysis has been concerned with the degradation of the power of the test during the degeneration of the model from the parametric case with θ_3 known to the parametric case with θ_3 unknown to the non-parametric case. The behavior of the power function of the test statistic λ_j for the one shot case is illustrated in Fig. 3. The power function of the hyperplane decision boundry is included for purposes of comparision. The test statistic for the parametric cases, (a) and (b), is computed on the basis of a two-sided composite alternative, whereas the sign test can only be computed for a one-sided composite alternative. Since the power associated with a one-sided alterative is generally higher than that for a two-sided alternative when tests are about the same parameter, the plot of the power function for the non-parametric case in Fig. 3 exceeds the corresponding plot for the parametric case with θ_3 unknown for certain values of $\delta/\sqrt{\theta_3}$. The consequent degradation of probability of error of the second kind can be extracted from Fig. 3 as

$$\beta = 1 - \beta(\delta/\sqrt{\theta_3}) \tag{4}$$

with the probability of error of the first kind equal to the signficance level α.

The extension of a dichotomization strategy from a "one shot" form to a recursive form is essential in order to effect the property of cognition or learning in signal processing. The concepts of classical sequential decision theory are of practically no value as developed, since the only intermediate decisions made therein are whether or not to continue observations of the scalar or vector samples. In order to realize the property of cognition or learning, it is:

(1) Intuitively felt, for the category model currently being investigated, that successive dichotomization on the basis of statistical similarity about a location parameter $\underline{\theta}$ with a reduction in the uncertainity about the dispersion or covariance matrix parameter Θ should give rise to an increase in the power of the test for given actual parameters with an increase in the length of the sequence observed (see hypothetical dashed line in Fig. 3).

(2) Necessary that the recursive form of the dichotomizing test statistic reflect the reduction in uncertainty about the parameters $(\underline{\theta}, \Theta)$.

In order to introduce the recursion based upon past observations into the current test statistic form, it is necessary to assume an abstract prior sequence $\underline{Z}_o$. Then for the assumption of a uniform prior kernel, $h(\{\theta\})$, in (3), the posterior kernel is of the Wishart form[3]. For an assumption of a Wishart kernel $h(\{\theta\})$, the posterior form is the multivariate form of the Pearson Type VII,

$$P(\underline{z}_1|\underline{Z}_o) = f(n,N_o,S_{N_o}) \left(1+\frac{1}{N_o} Q_{10}\right)^{-\frac{N_o}{2}} \tag{5}$$

where

$$f(n,N_o,S_{N_o}) = \frac{(2\pi)^{-\frac{n}{2}}\,\Gamma\left(\frac{N_o}{2}\right)}{\Gamma\left(\frac{N_o-n}{2}\right)\left(\frac{N_o}{2}\right)^{\frac{n}{2}}}$$

$$Q_{i0} = (\underline{z}_1 - \underline{\theta}_i)' S_{N_o}^{-1} (\underline{z}_1 - \underline{\theta}_i) \quad i = 1,2$$

$\mathrm{Ln}\, L(\omega_1 \cap \omega_2)$ then becomes (6)

$$\mathrm{Ln}\, L(\omega_1 \cap \omega_2) = -n\mathrm{Ln}2\pi + \mathrm{Ln} f(n,N_o,S_{N_o}) - \mathrm{Ln}|S_{N_o}| - \frac{N_o}{2}\left(\mathrm{Ln}\left[1 + \frac{Q_{10}}{N_o}\right] + \mathrm{Ln}\left[1 + \frac{Q_{20}}{N_o}\right]\right)$$

To obtain Max L, it is necessary to extend the abstract sequence, $\underline{Z}_o$, over a negative half line in order to obtain the asymptotic form of the Pearson Type VII kernel.

Thus

$$\lim_{N_o \to -\infty} \left(-\frac{No}{2}\right) LN \left(1 - \frac{(-Q_{io})}{N_o}\right) = -\frac{1}{2}\frac{Q_{io}}{2} \qquad i = 1,2 \tag{7}$$

$$\lim_{N_o \to -\infty} Ln\, f(n, N_o, S_{N_o}) = 0$$

and by proper factoring and combining,

$$LnL_{ASY}(\omega_1 \cap \omega_2) = -nLn2\pi - Ln|S_o + \Theta| - \frac{1}{4} TR\, S_j(S_o + \Theta) - \frac{1}{2} TR\, (\underline{\bar{z}} - \underline{\theta}_1)(\underline{\bar{z}} - \underline{\theta}_1)'(S_o + \Theta) \tag{8}$$

where

$$S_j = \sum_{k=1}^{j} (\underline{z}_k - \underline{\theta}_1)(\underline{z}_k - \underline{\theta}_1)' \qquad j = 2 \text{ for first two samples}$$

$$\text{Now } \frac{\partial LnL_{ASY}(\omega_1 \cap \omega_2)}{\partial \underline{\theta}_1} \Longrightarrow \underline{\theta}_{1_M} = \frac{\underline{z}_1 + \underline{z}_2}{2} = \underline{\bar{z}} \tag{9}$$

and for $(S_o + \Theta)$ positive semi-definite and S_j positive definite, the maximum of Ln $L_{ASY}(\omega_1 \cap \omega_2)$ with respect to the matrix parameter Θ occurs for

$$\Theta = 2\, S_j^{-1} - S_o \tag{10}$$

yielding $$\max_{\underline{\theta}, \Theta} Ln_{ASY}L\, (\omega_1 \cap \omega_2) = \frac{1}{(e\pi)^n |S_j|} \tag{11}$$

A similar abstraction, limiting process and factoring for $(\underline{\theta}, \Theta)$ ranging over the unrestricted space Ω yields

$$\max_{\underline{\theta}_1, \underline{\theta}_2, \Theta} Ln_{ASY}L(\Omega) = \frac{1}{(e\pi)^n |S_{A_1} + S_{A_2} + 2\underline{z}_o(\underline{z}_1 + \underline{z}_2)'|} \tag{12}$$

where $\underline{z}_o$ is a residual bias term as a result of the required factoring of $LnL_{ASY}(\Omega)$.

The test statistic for the "dichotomization" of the first two observations, $\underline{z}_1, \underline{z}_2$, is thus

$$\lambda_j(\underline{z}_1, \underline{z}_2) = \frac{|S_{A_1} + S_{A_2} + 2\underline{z}_o(\underline{z}_1 + \underline{z}_2)'|}{|S_j|} \qquad j = 1,2,\ldots,N \tag{13}$$

where S_j is given by (3) and S_{A_1}, S_{A_2} are individual auto-covariance matrices.

The steps necessary to incorporate the information about the parameters $\underline{\theta}, \Theta$ gained during sucessive observations have thus been developed for the initial observations. At this point, it is only intuitively concluded that successive tests for dissimilarity, dichotomization, will result in an increased power, $\beta(\underline{\delta}/\sqrt{\theta_3})$, with successive observations reinforcing the parameter S_j. It must be pointed out that if a UMP test statistic exists for the intended dichotomization, it is given by λ_j, however only for the "class of tests" based upon a test for dissimilarity about a location parameter $\underline{\theta}$ with a common dispersion matrix Θ for the two categories or populations.

Further work will include extensions to the m^{th} observation, refinements and analyses of boundedness and convergence, tests about other parameters and generalizations of the category models in an effort to establish the form of a recursive decision or operator structure required of cognition or "learning" without supervision or "teacher".

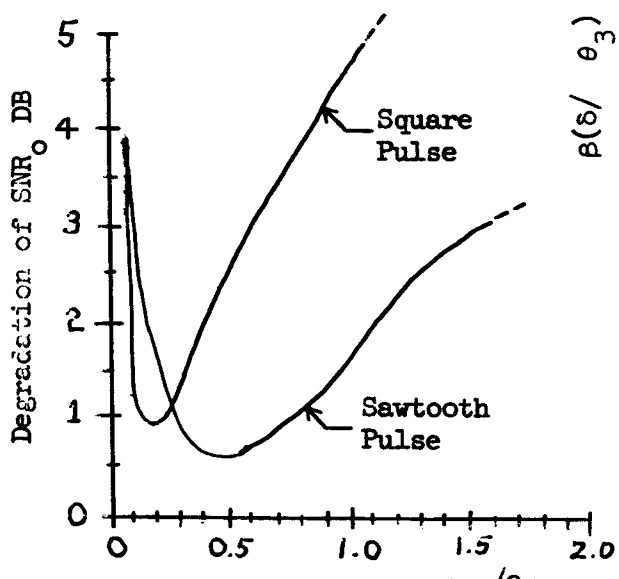

Fig. 2

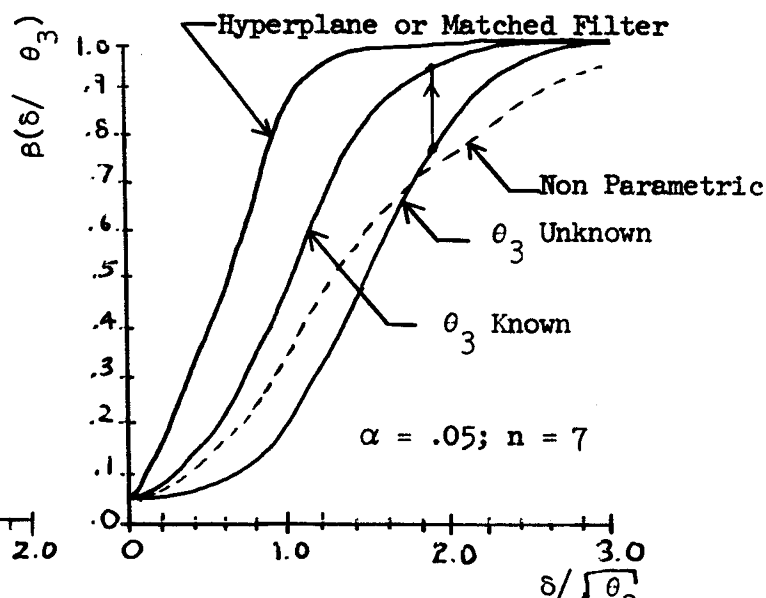

Fig. 3

REFERENCES

1. Raiffa, H. and Schlaifer, R. Applied Statistical Decision Theory Harvard University Press, Cambridge, Mass., 196_.

2. Wilks, S. S., Mathematical Statistics, J. Wiley and Sons, New York, N.Y., 1962.

3. Geisser, Seymour, "Bayesian Estimation in Multivariate Analysis," Annals of Mathematical Statistics, Vol. 36, No. 1-Feb. 1965.

4. Anderson, T. W., An Introduction to Multivariate Statistical Analysis, J. Wiley and Sons, New York, N.Y., 1958.

5. Fralick, "Learning to Recognize Patterns Without a Teacher," Stanford Electronics Laboratory, Report No. SU-SEL-65-011, April, 1965.

6. Hancock, J. C. and Gregg, W. D., "Cognitive Signal Processing," Purdue Univ. School of Electrical Engineering, First Semi Annual Research Summary, July-Dec. 1964, pp. 15-19, Section 5.1.

II. ADAPTIVE SYSTEMS

A. GENERALIZED KINEPLEX

P. A. Wintz

It is well known that the optimum (minimum probability of error) detector of known antipodal signals $s(t)$, $-s(t)$, $o \leq t \leq T$ in white gaussian noise $n(t)$ correlates the received data $x(t) = \pm s(t) + n(t)$ with a stored replica of the signal, and announces a decision in accordance with the sign of the correlator output at $t = T$. When the signal waveform is not known a priori, it may be reasonable to design a receiver capable of learning the unknown signal waveform and correlating the received data with the learned reference signal $r(t)$. If the reference waveform is constructed from K past bauds according to the decision-directed measurement strategy (see the Annual Report of Research, January 1965, p. 20) the system error rate is given by

$$\text{Prob}[d_o < 0|s(t)]\ \text{Prob}\ [s(t)] + \text{Prob}\ [d_o > 0|-s(t)]\ \text{Prob}\ [-s(t)]$$

where:

$$r_o(t) = x_1(t) + \sum_{i=1}^{K} \prod_{j=1} \text{sgn}\ (d_j)\ x_i(t)\ ;$$

$$d_j = \int_{-jT}^{(-j+1)T} x_j(t)\ r_j(t)\ dt.$$

In Fig. 1 shown below, $x_i(t)$ and $r_i(t)$ represent the data and reference waveforms during the i-th signaling interval.

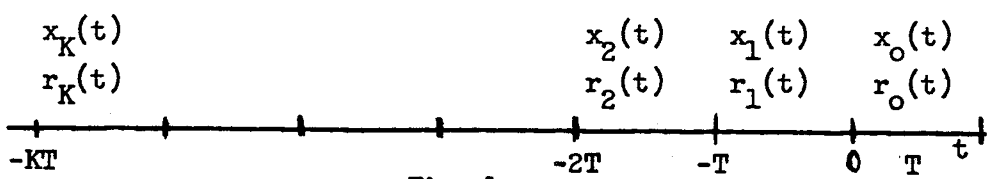

Fig. 1

By sampling the waveforms $x_i(t)$ and $r_i(t)$ N times in each T-second interval, d_j can be approximated by the sum $d_j = \sum_{i=1}^{N} x_j(t_i)\, r_j(t_i)$.

This formulation is suitable for Monte Carlo simulation on a digital computer. The results of this study are given in Fig. 2 where we have graphed the probability of error P_E as signal-to-noise ratio $R = \sum_{i=1}^{N} s^2(t_i)/2\, \overline{n^2(t_i)}$ for various values of measurement times K and numbers of samples N.

An analytical analysis of this problem is also being attempted. Severe mathematical difficulties are encountered since the decision-directed measurement strategy is inherently nonlinear. Therefore, for measurement times greater than unity (K > 1), the noise associated with the reference signal is nongaussian. Another problem arises because of the noise-cross-noise term associated with correlators using noisy references signals. It has been shown that for N = 2,4...

$$\text{Prob}\left[\sum_{i=1}^{N}(s_i + n_i)(a\, s_i + m_i) \le 0\right]$$

$$= \frac{1}{2} - \exp[-1(1+a^2)E/2] \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \frac{(\frac{aE}{2})^{j+1}[(1+a^2)E/8]^k}{(j+1)!k!} \sum_{n=0}^{j} \binom{j+2k+N+4+n}{j+k+\frac{N}{2}\; 2}\binom{j}{n}(-1)^n$$

where $E = \sum_{i=1}^{N} s_i^2$, and n_i and m_i are independent zero mean unit variance gaussian random variables.

Finally, an interesting identity was discovered in the course of the theoretical analysis. This identity, $\sum_{m=0} \binom{\alpha+m}{m} 2^{-(\alpha+m)} = 1$, $\alpha = 0,1,2,\ldots$ to the author's knowledge, has not been noted previously.

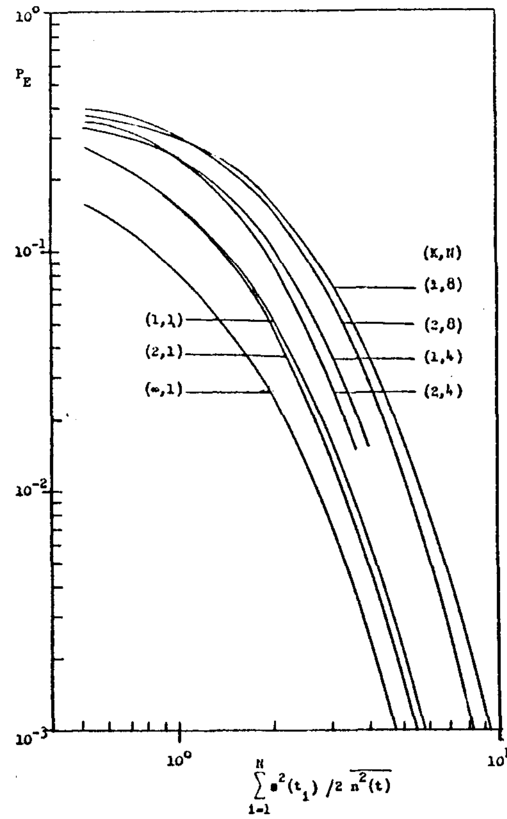

Fig. 2. Probability of Error vs Received Signal-To-Noise Ratio for Various Values of Measurement Time K and Numbers of Samples N.

B. SELF-SYNCRONIZING RECEIVERS

P. A. Wintz

E. J. Luecke

Investigation of digital communication systems under a variety of conditions and situations has shown that correlation techniques are required to achieve optimum performance. To implement a correlation receiver, it is necessary to know the epoch of each signaling period and the wave shape of the members of the signal set used in the system. For optimum results, correlation should be done at band pass. To generalize the problem, it is convenient to consider the frequency and phase of the "carrier" as well as the low pass wave shapes of the signal set as part of the signal waveshape. The problem of supplying to the receiver these parameters is not trivial under the reasonable constraints of maximum power, bandwidth, and signaling rate.

In one sense, the sub-optimality of a digital system is determined by the way in which the designer of the system desensitizes the receiver to the parameters of epoch and waveshape. To obtain optimum detection, the receiver must have these parameters. It is generally not possible to build into the receiver all of these parameters. Usually the only available parameter is the functional form of the low pass signal set. Thus, it is necessary for the transmitter to supply to the receiver the remaining necessary parameters. This process of bit synchronization is usually achieved by transmitting extra signals at the cost of total power and bandwidth.

The objective of the present investigation is to determine if measurements taken on the received signal will permit the determination of the epoch of the signal bits. If this is possible, the energy and bandwidth which is saved can be used for the improvement of the signal to noise ratio or the signaling rate. To simplify experimental procedures, the band pass case is not considered. Instead, the assumption is made that some demodulation scheme has supplied the low pass signal and additive noise.

Consider the demodulated signal $s(t) = \sum_{i=\infty}^{+\infty} S_i^m(\xi)$ where $S_i^m(\xi) = S_0^m\,(t - i\,T)$ for $i\,T < t \leq (i + 1)\,T$ and where $S^m(\lambda)$, $m = 1,2, K$ is the m^{th} signal of the k signals in the signal set. Assume that there is a probability distribution on "m" and the signals are transmitted independently.

Apply this $s(t)$ to a linear system which has impulse response $h(t)$. The output of the linear system can be written as

$$\lambda(t) = \int_0^{t-iT} S_i^m(\xi)h(t-i\,T-\xi)d\xi + \sum_{d=1}^{\infty} S_{i-d}^m(\xi)h(t-[i - d]\,T - \xi)\,d\xi$$

where $i\,T \leq t \leq [i+1]\,T$.

The desired response is any one which gives an indication when $t = (i + 1)\,T$. A little reflection on the form of $t(t)$, remembering that $S^m(\xi)$ is randomly chosen from the signal set, indicates that no linear operator will do the job.

If, however, some operation can be performed on the $s(t)$ before application to the linear system so that each $\bar{S}_{i-d}^m = f(S_{i-d}^m)$ is identical, then a linear system can be implemented to obtain epoch information. For the special case of binary antipodal signalling, the class of operators $y = f(\cdot)$ is given by all even functions.

The periodic signal which results from this "even function" operation may be determined by Fourier Series analysis. The zero crossings of the fundamental component of this signal are analytically related to the epoch of the bit. Bandpass filtering at the fundamental frequency will provide the desired epoch information.

The addition of noise to the input signal causes severe complications to the analysis of the system. At high S/N, the design objective would be to determine the combination of signal wave shape and nonlinear operator to provide maximum power at the fundamental frequency. At low S/N, the effect of the interaction of signal and noise in the nonlinear operator is not, in general, known. Under this mode of operation, the design objective would be to maximize the ratio of signal power at the fundamental to noise in the band around the fundamental. It is not apparent that the low S/N solution will be the same as the high S/N.

Because of the lack of analytic and experimental results on the distributions and spectral densities which result from nonlinear operations, the present point in this project is obtaining data for a number of representative signals with various S/N and representative nonlinear operators.

C. ADAPTIVE PROCESSING OF TROPO-SCATTER DATA

P. A. Wintz

M. D. Shapiro

In this experiment, certain of the concepts discussed in Part A of this section are being used to process binary data transmitted over the Purdue-Collins Radio Co. tropospheric-scatter link. The received data are first time sampled and the samples stored on magnetic tape. These data are then processed in various ways on Purdue's 7094 Computer. Computer programs have been written for the following detection strategies:

1. The binary information is differentially encoded, and the received data are

correlated with a reference signal. For a measurement time of one signal duration, the reference signal is simply the data received during the previous baud. For measurement times greater than one signal duration, the data from K past bauds are used to form a reference signal according to the decision-directed measurement strategy.

2. This process is similar to that in 1, except that a nondecision-directed measurement strategy is used to form a reference from the K past bauds. Each sample of the references is taken to be the sample mean of the magnitude of the corresponding samples of the previous K bauds.

3. In this process the received data are correlated with stored replicas of the transmitted signals.

These programs have been debugged and used to process data originated in the laboratory. Processing of actual tropo data will start in September, 1965. All processing is being done at baseband.

III. SIGNAL DESIGN

A. TROPO-SCATTER SIGNAL DESIGN

D. R. Anderson

In the design of signals for a variable communication channel, the central quantity is the ambiquity function which is defined for any pair of signals $s_1(t)$ and $s_2(t)$ by the formula

$$X_{1,2}(\omega,\tau) = \int_{-\infty}^{+\infty} s_1(t)\ s_2(t+\tau)e^{j\omega t}dt \tag{1}$$

Although $X_{1,2}(\omega,\tau)$ originally arose in the analysis of radar observations of a fixed object by a matched filter, Price and Green[1] showed the importance of it in multipath communication in their analysis of the Rake system. The same authors have since pointed out the importance of $|X_{1,2}(\omega,\tau)|^2$ for scattering and multipath channels even when optimum detection does not require matched filters. They have shown that figures of merit for a signal set are 1) rectangular white-noise-like spectrum for every member, 2) uniformly small ambiguity functions for every pair of members, 3) a sharply peaked self-ambiguity function for each individual member.

For any pair of equal-duration signals one can find a basic lower bound in terms of time-bandwidth products for the time r.m.s. value of their cross-ambiguity function. If we call an arbitrary pair of signals $s_1(t)$ and $s_2(t)$, their common time-duration T, and their bandwidths B_1 and B_2, the bound is given by:

$$\max_{\omega} \left[\frac{1}{T} \int_{-T}^{T} |X_{1,2}(\omega,\tau)|^2 dt \right]^{1/2} \geq \frac{1}{2} / \left(2TB_1 + 2TB_2\right)^{1/2} \qquad (2)$$

This shows in particular that we have:

$$\max_{\omega} \left(\max_{\tau} |X_{1,2}(\omega,\tau)|\right) \geq \frac{1}{2}/(2TB_1 + 2TB_2)^{1/2} \qquad (3)$$

That is to say, the global maximum of $|X_{1,2}(\omega,\tau)|$ must be at least $\frac{1}{2}/(2TB_1 + 2TB_2)^{1/2}$. Since one can construct realizable pairs of signals for which $|X_{1,2}(\omega,\tau)|$ is no more than $5/(2TB_1+2TB_2)^{1/2}$, (2) gives fundamental information about how small a crossambiguity function can be.

REFERENCES

1. Price, R. and P. Green, "Signal Processing in Radar Astronomy Communication via Fluctuating Multipath Media," M.I.T. Lincoln Lab., Lexington, Mass., Tech, Rept. No. 234; October, 1960.

B. DIGITAL COMMUNICATION SYSTEMS OPTIMIZATION FOR CHANNELS WITH MEMORY

J. C. Hancock

E. A. Quincy

1. Re-Statement of the Problem

The specific problem considered in this research is the optimization of entire binary communication systems, i.e., joint optimization of the transmitted pulse waveforms and the receiver when the channel response is time-invariant and known. Also, the channel is assumed to exhibit sufficient memory such that intersymbol interference results at the receiver. The criterion of optimality considered is minimum average probability of detection error. For recent literature pertinent to this problem see Refs. 1,2,3, and 4.

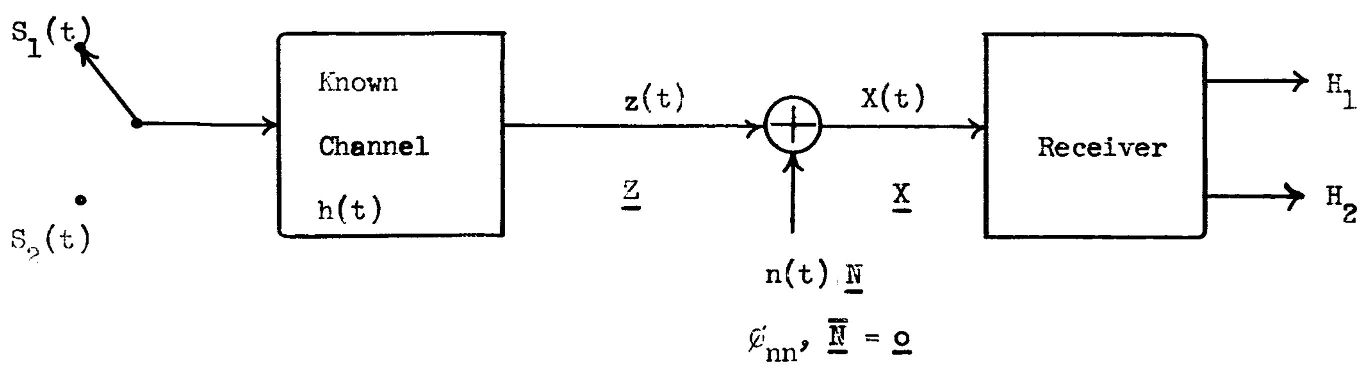

Fig. 1 Binary Communication System Model

Figure 1 shows a model of the binary communication system considered in this research. The additive noise is assumed to be gaussian with zero mean (AGNZM) and covariance ϕ_{nn}. Also, the received signal is assumed to be representable by a finite sum of weighted basis functions such that the weighting coefficients are a k-dimensional vector denoted by a bar beneath an upper-case letter.

2. Initial Approach

An ideal approach to this problem is to derive the Bayes receiver from the Maximum Likelihood Ratio, $\lambda(X)$, for arbitrary transmitted signal waveforms, $s_1(t)$ and $s_2(t)$, of length T and an arbitrary time-invariant channel with impulse response $h(t)$. Then ideally, the average probability of error P_e would be derived for this receiver. The resulting P_e would be a function of the received signal energies $\underline{E}$, of all possible cross-correlations of the two desired signal wave forms with all possible combinations of received sequences of overlapping pulses, and of the noise covariance matrix ϕ_{nn} and the known impulse response $h(t)$. Hence for a specific impulse response and noise we can denote

$$P_e = f(E, \underline{\rho}) = g\left[s_1(t),\ s_2(t)\right] \tag{1}$$

With an explicit expression available for P_e it would then be a matter of applying variational techniques to minimize P_e with respect to the transmitted waveforms with constraints that fix the input energy to the channel, i.e.,

$$\text{MIN} \left\{P_e = g\left[s_1(t),\ s_2(t)\right]\right\} \tag{2a}$$

with

$$\int_{-T/2}^{T/2} s_1{}^2(t)\ dt = E_1 \tag{2b}$$

$$\int_{-T/2}^{T/2} s_2^2(t)\ dt = E_2 \tag{2c}$$

For a received signal $\underline{Z}$ and AGNZM noise N, let the received waveform be

$$\underline{X} = \underline{Z} + \underline{N} \tag{3}$$

Then the corresponding Maximum Likelihood Ratio for M bauds of overlap of the received signals is

$$\lambda(\underline{x}) = \frac{P(\underline{X}|s_1)}{P(\underline{X}|s_2)} = \frac{\sum_{i=1}^{r} P_i \, P(\underline{X}|s_1, \underline{Z}_{1i})}{\sum_{i=1}^{r} P_i \, P(\underline{X}|s_2, \underline{Z}_{2i})} \tag{4}$$

where the receiver observation period is $[0 - (M + 1)T]$ and

$$r = 2^{2M} \tag{5}$$

is the number of possible waveforms that could be received on this observation period given that s_i was sent at the beginning of the observation period. For the AGNZM noise considered (4) becomes

$$\lambda(\underline{x}) = \frac{\sum_{i=1}^{r} \frac{P_i}{(2\pi)^{K/2}|\phi_{nn}|^{1/2}} \epsilon^{-\frac{1}{2}(\underline{X} - \underline{Z}_{1i})^T \phi_{nn}^{-1} (\underline{X} - \underline{Z}_{1i})}}{\sum_{i=1}^{r} \frac{P_i}{(2\pi)^{K/2}|\phi_{nn}|^{1/2}} \epsilon^{-\frac{1}{2}(\underline{X} - \underline{Z}_{2i})^T \phi_{nn}^{-1} (\underline{X} - \underline{Z}_{2i})}} \tag{6}$$

$$= \frac{\sum_{i=1}^{r} P_i \quad \underline{X}^T \phi_{nn}^{-1} \underline{Z}_{1i} - \frac{1}{2} \underline{Z}_{1i} \phi_{nn}^{-1} \underline{Z}_{1i}}{\sum_{i=1}^{r} P_i \quad \underline{X}^T \phi_{nn}^{-1} \underline{Z}_{2i} - \frac{1}{2} \underline{Z}_{2i} \phi_{nn}^{-1} \underline{Z}_{2i}} \tag{6a}$$

The first term of the exponent, in both numerator and denominator, of (6a) represents cross-correlation of the received waveform with one of the possible waveforms weighted by the noise covariance. The second term is a form of energy to noise ratio for this particular waveform. In general the receiver corresponding to the likelihood ratio (6a) is very non-linear and complicated. Even though the receiver could be implemented in its complicated form, the analysis of the performance is non-tractable due to the non-linearities. Hence the expression for probability of error,

$$P_e = g\left[s_1(t),\ s_2(t)\right] \tag{7}$$

desired in order to have an overall system functional to minimize, cannot be obtained unless further simplifying assumptions are made.

3. Alternate Approaches and Results

Since the above approach leads to a formidable problem, i.e., obtaining the functional (7), other approaches were pursued which circumvent this problem.

3.1 Special Case (Linear Bayes Receiver)

Upon investigation of

$$\lambda'(\underline{X}) = \log \lambda(\underline{X}) \tag{8}$$

in (6a), there are special cases of interest which reduce $\lambda'(\underline{X})$ to a linear form. For example, the case of an RC-low-pass channel with adjacent baud overlap of received signals (M = 1) and equally-probable, bipolar signals s_1, s_2. In addition if the received waveforms on (-T/2 to 3T/2)

$$\underline{Z}_{1i},\ \underline{Z}_{2i} \quad ; \quad i = 1,2,3,4 \tag{9}$$

are all equal energy then $\lambda'(\underline{X})$ was shown to reduce to the linear form

$$\lambda'(\underline{X}) = 2\ \underline{X}^T \phi_{nn}^{-1} \underline{Z}_1 \tag{10}$$

where $\underline{Z}_1$ is the channel output corresponding to a single-shot s_1 input. Also upon investigation of the equal energy sequences condition, it was shown, by Quincy[5], to occur when the head and tail of $\underline{Z}_1$ were orthogonal. Hence in order for this linear form (10) to be a Bayes receiver, the signal s_1 should be designed to maximize energy transfer through the channel with the constraint that the head and tail out of the channel are orthogonal. For this bipolar case

$$s_1(t) = -\ s_2(t) \tag{11}$$

The average probability of error was derived for this case of intersymbol interference with the additional assumption that the noise was white with spectral density N_o and band-limited to the signal bandwidth. The resulting probability of error was

$$P_e = \Phi\left(-\sqrt{\frac{E}{N_o}}\right) \tag{12}$$

where

$$E \triangleq \int_{-T/2}^{3T/2} z_1^2(t)\, dt \tag{13}$$

and

$$\Phi(-t) \triangleq \int_{-\infty}^{t} e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} \tag{14}$$

Variational techniques have not thus far yielded a solution for s_1 above. However, Chalk[6] showed that the signal which maximizes energy transfer for the RC channel without an additional constraint is

$$s_1(t) = A \operatorname{Cos} \omega_o t, \quad -\frac{T}{2} \leq t \leq T/2 \tag{15}$$

and if this is shifted in phase by a specified amount, i.e.,

$$s_1(t) = A_1 \operatorname{Cos} (\omega_o t + \theta), \quad -T/2 \leq t \leq T/2 \tag{16}$$

then the head and tail will be orthogonal. Unfortunately the energy transfer is reduced as θ is changed from zero. This procedure yields a simple linear Bayes receiver for a special case of intersymbol interference. However it does not yield an overall optimum system, only an optimum receiver for these signals. If the optimum signal can be found which maximizes energy transfer with the constraint that the head and tail are uncorrelated, it will be interesting to note how much energy transfer is decreased by this constraint. From experience we would expect it to be decreased slightly. However, if the optimum signal transfers the same energy as (15) while meeting the additional constraint, then an overall optimum system will be obtained.

3.2 Approximate Likelihood Ratio $\tilde{\lambda}(\underline{X})$

For the case of equally-probable signals s_1, s_2, (6a) has an interesting linear approximation. Let (6a) be written as

$$\lambda(\underline{X}) = \frac{\sum_i e^{f_i(\underline{X})}}{\sum_i e^{g_i(\underline{X})}} \simeq e^{\sum_i \left[f_i(\underline{X}) - g_i(\underline{X}) \right]}$$

$$\tilde{\lambda}(\underline{X}) = e^{\sum_i \left[f_i(\underline{X}) - g_i(\underline{X}) \right]} = e^{\sum_{i=1}^{r} \left[\underline{X}^T \phi_{nn}^{-1}(\underline{Z}_{1i} - \underline{Z}_{2i}) - \frac{1}{2}(\underline{Z}_{1i}\phi_{nn}^{-1}\underline{Z}_{1i} - \underline{Z}_{2i}\phi_{nn}^{-1}\underline{Z}_{2i}) \right]} \tag{18}$$

and

$$\tilde{\lambda}'(\underline{X}) = \log \tilde{\lambda}(\underline{X}) \tag{19}$$

$$= \underline{X}^T \sum_{i=1}^{r} \phi_{nn}^{-1}(\underline{Z}_{1i} - \underline{Z}_{2i}) - \frac{1}{2} \sum_{i=1}^{r} (\underline{Z}_{1i}\phi_{nn}^{-1}\underline{Z}_{1i} - \underline{Z}_{21}\phi_{nn}^{-1}\underline{Z}_{2i}) \tag{19a}$$

Since (19a) is a linear operation on $\underline{X}$ the receiver performance in terms of P_e can readily be derived and variational techniques applied to minimize P_e with respect to s_1, s_2.

It is interesting to note that $\tilde{\lambda}(\underline{X})$ in (18) is a consistent approximation as the intersymbol interference is reduced to zero, i.e., for this case M = 0, r = 1 and

$$\tilde{\lambda}(\underline{X}) = \lambda(\underline{X}) \tag{20}$$

Also, log $\tilde{\lambda}(\underline{X})$ in (19) reduces exactly to that of (9) for the same case considered there.

The average probability of error was derived for a special case of (19a) i.e., for white noise of spectral density N_o, band-limited to the signal bandwidth, adjacent baud overlap only ($M = 1$, $r = 4$) and a low-pass RC channel. The resulting probability of error corresponding to the receiver implementing (19a) is

$$P_e = \frac{1}{2}\,\Phi\left[-\sqrt{\frac{E}{N_o}}\,\right] + \frac{1}{4}\,\Phi\left[-\sqrt{\frac{E}{N_o}}\,(1 - \rho_{ht})\,\right] + \frac{1}{4}\,\Phi\left[-\,\frac{E}{N_o}\,(1 + \rho_{ht})\,\right] \tag{21}$$

where E is given by (13) and Φ by (14). The normalized head-tail correlation is given by

$$\rho_{ht} = \frac{\int_{-T/2}^{T/2} z_1(t)\, z_1(t + T)\, dt}{E} \tag{22}$$

and

$$z_1(t) = \int_{-T/2}^{t} s_1(\tau)\, h(t - \tau)\, d\tau \tag{23}$$

Hence (21) can now be written as a function of s_1 and s_2 as in (7)

$$P_e = g\left[s_1(t),\ s_2(t)\right] = g\left[s_1(t),\ -\ s_1(t)\right] \tag{24}$$

The ideal step at this point is to apply variational techniques to (21) to find the optimum waveform s_1 which minimizes P_e with the constraint that the energy into the channel is fixed. We realize of course that the resulting system would not necessarily be the overall optimum system but could be; i.e., it is an upper bound on the overall optimum system

$$P_e^o \leq P_e^a \tag{25}$$

Also this probability of error converges to the optimum system probability of error as ρ_{ht} tends toward zero, i.e.,

$$P_e^a \xrightarrow[\rho_{\Omega t} \to 0]{} P_e^o \tag{26}$$

as can be seen by comparing (12) and (21). Actually, the approximate likelihood ratio converges to the exact one in this case. Even though we cannot derive the analytical expression for P_e^o to minimize by signal design, we note from (25) that we will force it smaller and smaller as we do minimize P_e^a. Hence, if $P_e^o = P_e^a$ for that case we have the jointly optimum transmitter and receiver. If not, we can compare P_e^a against P_e^o obtained by applying Monte Carlo techniques to the exact likelihood ratio. Hopefully this will show that P_e^a is sufficiently close to P_e^o for practical considerations.

In order to gain further insight as to the trade-off involved between E and ρ_{ht} and their effect on P_e^a an ad-hoc approach was temporarily ensued at this point. The signal which maximizes energy transfer for this channel corresponding to (16) was derived for $\theta = 0$, i.e.,

$$S_1(t) = A_1 \cos(\omega_o t + \theta), \quad -\frac{T}{2} \leq t \leq \frac{T}{2} \tag{27}$$

Next θ was allowed to vary and $A_1(\theta)$ was computed for each θ since the energy input was held fixed. Now the energy output E and $\rho_{\Omega t}$ were computed as a function of θ, i.e.,

$$E = f_1(\theta) \tag{28}$$

$$\rho_{\Omega t} = f_2(\theta) \tag{29}$$

and eventually the probability of error was determined as a function of θ

$$P_e^a = g(\theta) \tag{30}$$

By allowing θ to vary, the minimum in P_e^a was found which corresponded to $\theta = -49.3^o$. Figure 2 shows a sketch of this performance compared to the Bayes receiver evaluated by

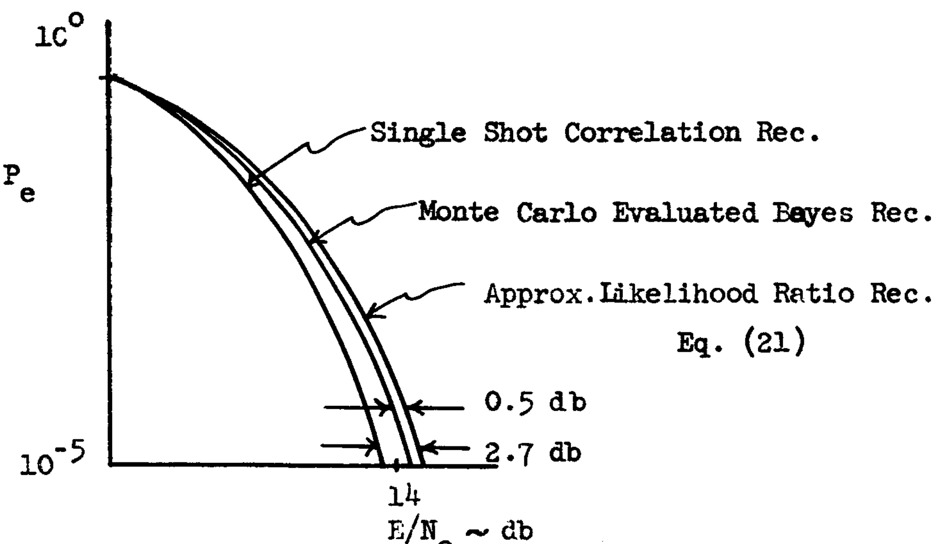

Fig. 2 Sketch of Relative Performance

Monte-Carlo techniques and to the single-shot case of correlation reception: Each system had the same signal applied. It is especially interesting to note that the Bayes receiver is only about one-half db better in performance at high signal-to-noise ratio.

REFERENCES

1. Quincy, E. A., 1st Quarterly Progress Report "Digital Communication Systems Optimization for Channels with Memory", Department of Electrical Engineering (To J. C. Hancock), Purdue University, December 1964.

2. Tufts, D. W. and Aron, J. R., "Intersymbol Interference and Estimation of Digital Message Sequences", International Conf. on Microwaves, Circuit Theory, Summaries of Papers, Part 3, Information Theory, pp. 55-56, Tokyo, Japan, September, 1964.

3. Tufts, D. W., "Nyquist's Problem - The Joint Optimization of Transmitter and Receiver in Pulse Amplitude Modulation," Proc. of IEEE; March, 1965, pp: 248-259.

4. Tufts, D. W., and Shnidman, D. A., "Effect of Correlated Data and Colored Noise on Error Probability and Mean Square Error in PAM data Transmission," IEEE Int'l. Conv. Record, Part 7; March, 1965, pp. 227-237.

5. Quincy, E. A., Ph.D. Thesis Proposal "Joint Optimization of Transmitter and Receiver in Binary Communication Systems with Channel Memory", Department of Electrical Engineering, Purdue University, March, 1965.

6. Chalk, J., "The Optimum Pulse-Shape for Pulse Communication," Proc. IEE Vol. 97, Part III: 88-02, Savoy Place, London W.C.2, England, March, 1950.